CoAuthor: Designing a Human-AI Collaborative Writing Dataset for Exploring Language Model Capabilities

Year: 2022 | *Citations: 438* | *Authors: Mina Lee, Percy Liang, Qian Yang*

Abstract

Large language models (LMs) offer unprecedented language generation capabilities and exciting opportunities for interaction design. However, their highly context-dependent capabilities are difficult to grasp and are often subjectively interpreted. In this paper, we argue that by curating and analyzing large interaction datasets, the HCI community can foster more incisive examinations of LMs' generative capabilities. Exemplifying this approach, we present CoAuthor, a dataset designed for revealing GPT-3's capabilities in assisting creative and argumentative writing. CoAuthor captures rich interactions between 63 writers and four instances of GPT-3 across 1445 writing sessions. We demonstrate that CoAuthor can address questions about GPT-3's language, ideation, and collaboration capabilities, and reveal its contribution as a writing "collaborator" under various definitions of good collaboration. Finally, we discuss how this work may facilitate a more principled discussion around LMs' promises and pitfalls in relation to interaction design. The dataset and an interface for replaying the writing sessions are publicly available at https://coauthor.stanford.edu.